Clinical Chemistry 48:10
1835–1843 (2002)

Oak Ridge Conference

Boosted Decision Tree Analysis of Surface-enhanced Laser Desorption/Ionization Mass Spectral Serum Profiles Discriminates Prostate Cancer from Noncancer Patients

Yinsheng Qu,[1] Bao-Ling Adam,[2,4] Yutaka Yasui,[1] Michael D. Ward,[2,4] Lisa H. Cazares,[2,4] Paul F. Schellhammer,[2,3,4] Ziding Feng,[1] O. John Semmes,[2,4] and George L. Wright, Jr.[2,3,4*]

Background: **The low specificity of the prostate-specific antigen (PSA) test makes it a poor biomarker for early detection of prostate cancer (PCA). Because single biomarkers most likely will not be found that are expressed by all genetic forms of PCA, we evaluated and developed a proteomic approach for the simultaneous detection and analysis of multiple proteins for the differentiation of PCA from noncancer patients.**

Methods: **Serum samples from 386 men [197 with PCA, 92 with benign prostatic hyperplasia (BPH), and 96 healthy individuals], randomly divided into training (n = 326) and test (n = 60) sets, were analyzed by surface-enhanced laser desorption/ionization (SELDI) mass spectrometry. The 124 peaks detected by computer analyses were analyzed in the training set by a boosting tree algorithm to develop a classifier for separating PCA from the noncancer groups. The classifier was then challenged with the test set (30 PCA samples, 15 BPH samples, 15 samples from healthy men) to determine the validity and accuracy of the classification system.**

Results: **Two classifiers were developed. The AdaBoost classifier completely separated the PCA from the noncancer samples, achieving 100% sensitivity and specificity. The second classifier, the Boosted Decision Stump Feature Selection classifier, was easier to interpret and used only 21 (compared with 74) peaks and a combination of 21 (vs 500) base classifiers to achieve a sensitivity and specificity of 97% for the test set.**

Conclusions: **The high sensitivity and specificity achieved in this study provides support of the potential for SELDI, coupled with a bioinformatics learning algorithm, to improve the early detection/diagnosis of PCA.**

The search for biomarkers for the early detection and diagnosis of cancer has been a daunting task with little success. Much of the effort in the past has largely centered on the discovery and characterization of single markers. On the basis of the marked microheterogeneity of most human cancers, it is doubtful that a single gene, chromosome aberration, or protein will be discovered that is expressed by all phenotypic forms of a cancer. For example, prostate-specific antigen (PSA)[5] is probably the best example of the use of a single biomarker as an aid in the diagnosis of prostate cancer (PCA). However, its low specificity in distinguishing PCA from benign prostatic hyperplasia (BPH) limits its use as an early detection biomarker, and preoperative serum values <10 μg/L are not useful for predicting either the presence of disease or postoperative outcome *(1)*. It is important that better biomarkers be identified to reach the goal of reducing the mortality of PCA. To reach this goal, rapid, high-through-

[1] Cancer Prevention Research Program, Fred Hutchinson Cancer Research Center, Seattle, WA 98109.

Departments of [2] Microbiology and Molecular Cell Biology and [3] Urology, Eastern Virginia Medical School, Norfolk, VA 23501.

[4] Virginia Prostate Center, Eastern Virginia Medical School and Sentara Cancer Institute, Norfolk, VA 23501.

*Address correspondence to this author at: Department of Microbiology and Molecular Cell Biology, Eastern Virginia Medical School, 700 West Olney Rd., Norfolk, VA 23507. Fax 757-624-2255; e-mail wrightgl@evms.edu.

Received May 6, 2002; accepted July 23, 2002.

[5] Nonstandard abbreviations: PSA, prostate-specific antigen; PCA, prostate cancer; BPH, benign prostatic hyperplasia; SELDI, surface-enhanced laser desorption/ionization; QC, quality control; PBS, phosphate-buffered saline; AUC, area under the curve; and BDSFS, Boosted Decision Stump Feature Selection.

put technology that can both detect and simultaneously analyze multiple biomarkers will be required.

A recent advance in mass spectrometry, surface-enhanced laser desorption/ionization (SELDI) time-of-flight mass spectrometry, provides a sensitive system to detect and resolve multiple proteins bound to protein chip arrays *(2, 3)*. Studies from our laboratory have demonstrated the successful use of SELDI to detect PSA and other known prostate biomarkers in body fluids and cell lysates, including the discovery of potential new biomarkers *(4)*. In addition to being a platform for biomarker discovery, the SELDI system can serve as a clinical assay format, offering a distinct advantage over current two-dimensional electrophoresis systems. With this platform, SELDI protein profiles, spectral patterns, or "fingerprints" from the test samples are compared with the spectrum of the control sample. The successful use of SELDI protein profiling to differentiate cancer from noncancer was first reported by our laboratory for bladder cancer *(5)*. In this study, differential analyses of the urine protein patterns increased the detection rate of early stage bladder cancer by 60% when compared with a diagnosis by urine cytology. Comparison of the spectral patterns was performed by manual visual inspection, a laborious undertaking fraught with significant error, and clearly suggested that bioinformatic classifier algorithms will be required to efficiently and effectively deal with the high dimensionality of the SELDI data.

A boosting decision tree algorithm was used in this study to analyze the n-dimensional SELDI data and develop a classifier for discriminating men with PCA from men with BPH and healthy age-matched men. A standardized training data set was used to construct a classifier that could completely discriminate men with PCA from men with BPH and healthy men when tested with a blinded test set. These results both demonstrate and support the clinical utility of the SELDI protein profiling system coupled with an artificial intelligent classifier as a potentially powerful and innovative proteomic assay for the early detection/diagnosis of PCA.

Materials and Methods

STUDY GROUPS AND SAMPLES

Serum samples were obtained from the Virginia Prostate Center Tissue and Body Fluid Bank. All samples had been procured from consenting patients according to protocols approved by the Institutional Review Board and stored frozen at −8 °C. None of the samples had been thawed more than twice. Age-matched pretreatment samples from 99 PCA patients diagnosed with organ-confined cancer, 98 PCA patients with non-organ-confined disease, 92 patients diagnosed with BPH (PSA <10 μg/L and negative biopsy), and specimens from 96 healthy men (negative digital rectal examination, PSA <4.0 μg/L, and no evidence of prostatic disease). The mean PSA values were 1.32 μg/L for healthy men, 4.60 μg/L for men with BPH, 10.10 μg/L for men with organ-confined PCA, and 206.93 μg/L for men with non-organ-confined PCA. A quality-control (QC) sample was prepared by pooling an equal amount of serum from each healthy donor and storing 100-μL aliquots at −8 °C.

EXPERIMENTAL DESIGN

Duplicate serum samples from healthy men (n = 96), men with BPH (n = 93), men with organ-confined PCA (n = 99), and men with non-organ-confined PCA (n = 98) were processed over a 2-week timeframe. A bioprocessor, which holds 12 chips in place, was used to process 96 samples at one time. Each chip contained two "QC spots", consisting of serum from one universal cancer patient and one normal pooled serum, which were applied to each chip along with the test samples in a random fashion. The QC spots served as quality control for assay and chip variability. The samples were blinded for the technicians who processed the samples. The overall data set size was determined based on the power analysis for validation sample size to obtain 95% confidence, 90% sensitivity, and 75% specificity.

The reproducibility of the SELDI spectra, i.e., mass and intensity from array to array on a single chip (intraassay) and between chips (interassay), was determined with the pooled normal serum QC sample. Seven proteins in the range of 3–10 kDa observed on spectra randomly selected over the course of the study were used to calculate the CV. The intra- and interassay CVs for mass were both 0.05%, and the intra- and interassay CVs for the normalized intensity were 15% and 20%, respectively.

SELDI PROTEIN PROFILING

Serum samples were prepared by vortex-mixing 20 μL of serum with 30 μL of 8 mol/L urea containing 10 mL/L CHAPS in phosphate-buffered saline (PBS) in a 1.5-mL microcentrifuge tube at 4 °C for 10 min. This was followed by the addition of 100 μL of 1 mol/L urea containing 1.25 mL/L CHAPS, and the mixture was briefly vortex-mixed. The samples were diluted 1:5 in PBS and applied to each well of a bioprocessor (Ciphergen Biosystems) containing IMAC-3 chips previously activated with $CuSO_4$. The bioprocessor was then sealed and agitated on a platform shaker at a speed of 250 rpm for 30 min. A pooled QC serum sample, prepared in the same manner, was applied to an array on each chip used in each experiment as a reproducibility control. The excess serum mixture was discarded, and the chips were washed three times with PBS. The chips were then removed from the bioprocessor, washed 10 times with deionized water, air-dried, and stored in the dark until subjected to SELDI analysis. Before SELDI analysis, 0.5 μL of a saturated solution of sinapinic acid in 500 mL/L acetonitrile containing 5 mL/L trifluoroacetic acid was applied onto each chip. The sinapinic acid was applied twice, and the array surface was allowed to air dry between each application. Chips were placed in the PBS-II mass spectrometer (Ciphergen), and time-of-flight spectra were generated by averaging

192 laser shots in positive mode with a laser intensity of 220, detector sensitivity of 7, and a focus lag time of 900 ns. Mass accuracy was calibrated externally using the All-in-1 peptide molecular weight standard (Ciphergen). Peak detection and alignment were performed with Ciphergen ProteinChip Software 3.0 with slight modifications. The mass range from 2 to 40 kDa was selected for analysis because this range contained the majority of the resolved protein/peptides.

DATA ANALYSIS

Feature selection. We use the peaks in the range from 2 to 40 kDa as predictors. The power of each peak in discriminating men with PCA from healthy men, men with BPH from healthy men, and men with BPH from men with PCA was determined by estimating the area under the ROC curve (AUC), which ranges from 0.5 (no discrimination) to 1.0 (absolute prediction) *(6)*. The peaks with an AUC <0.62 were excluded from further data analyses.

Boosted decision stump classifier. Boosting is used to reduce the error of any "weak" learning algorithm. A decision tree with only one split is called a decision stump, which usually is a weak learner. However, the AdaBoost algorithm described by Freund and Schapire *(7)*, Hastie et al. *(8)*, and Friedman et al. *(9)* can combine those weak learners into an accurate classifier. The combined classifier is a committee with the decision stumps, the base classifiers, as its members. The committee makes a decision by a majority vote. The base classifiers are constructed on weighted examples. For the first round, equal weights to all examples are used. For the next round, the weights are increased for the examples misclassified by the first decision stump and decreased for the examples correctly classified by the first decision stump. Therefore, the second decision stump focuses on the samples misclassified by the first stump. This procedure is repeated again and again until a defined number of stumps have been created. In the committee, each member has its own specialty arising from its special training. For example, the second member's specialty is to correct the first member's mistake, the third member's specialty is to correct the second member's mistake, and so forth. Therefore, the committee with diverse members can do a better job than any single member.

For a given sample, the classification decision is made by a majority vote by the base classifiers. If $V1$ is the total vote of the base classifiers who made correct decisions, and $V2$ is the total vote of the base classifiers who made wrong decisions, then if $V1 > V2$, the combined classifier will make a correct decision. The quantity $V1 - V2$ is the margin of the vote result. If the margin is positive, the sample is classified correctly. As the margin increases, the confidence becomes greater. Boosting reduces test error by increasing the minimal margin in the training set. This process is explained in greater detail in a data supplement that can be viewed with the online version of this article at http://www.clinchem.org/content/vol48/issue10/. It is also further elaborated on our website (http://140.107.129.65/stat_methods.htm).

To illustrate this method, we simulated a data set with 100 noncancer and 100 cancer samples. Each sample had records of activities for 12 peaks. Peaks 1, 2, and 3 were higher in noncancer samples, and peaks 4, 5, 6, and 7 were higher in cancer samples. Peaks 8, 9, 10, 11, and 12 were not informative (Fig. 1). We then ran the AdaBoost algorithm for five cycles. In each round, a peak was selected. The selected peaks in the five rounds were 4, 2, 1, 4, and 2. The first decision stump was: if peak 4 is >1.80, then classify as cancer. The classifier combined the five decision stumps and yielded a training error of 2% (4 of 200 were misclassified). We also simulated a test set from the same distribution and applied the combined classifier to the test set. The error rate was 5.5% (11 of 200 were misclassified).

Definition and evaluation of sensitivity and specificity. In this report, sensitivity is defined as the conditional probability of predicting cancer given that the gold standard is cancer. Likewise, we define specificity as the conditional probability of predicting noncancer given that the gold standard is noncancer. We use the Bayesian approach to calculate the estimated sensitivity and specificity with β prior [$\beta(a, b)$], where a and b are the two hyperparameters. If r of n cancer cases are predicted as cancer cases, then the expected sensitivity is $(r + a)/(n + a + b)$. We chose $a = b = 1$. If $r = n = 30$, then the expected sensitivity is $31/32 = 96.9\%$, and the 95% confidence interval can be determined from the posterior distribution β $(a + r, b + n - r)$. If the conventional likelihood approach is used, the sensitivity will be $31/31 = 100\%$ and the 95% confidence interval will be difficult to calculate.

Results

Each SELDI spectrum revealed an average of 80 peak masses in the 2–40 kDa range. The QC spectra were very reproducible with intra- and interassay CVs for peak location of 0.05%, and CVs of 15% and 20%, respectively, for peak intensity (data not shown). Fig. 2 shows a representative example of the SELDI spectra. Analysis of all 772 spectra (336 samples run in duplicate) identified 779 peaks, of which 124 had an AUC ≥ 0.62. These 124 peaks identified in the training set were used to construct the classifier.

One of the concerns in the construction and use of learning algorithms is the possibility of overfitting the data. However, boosting methods can avoid overfitting by increasing the minimal margin. The larger the minimal margin the less chance of a test sample being misclassified. Fig. 3 shows the minimal margin and the generalization error rate (testing error) against the number of base stumps for the boosted decision tree classifier distinguishing noncancer from cancer. After the training error reached zero (round 47), the minimal margin kept increas-

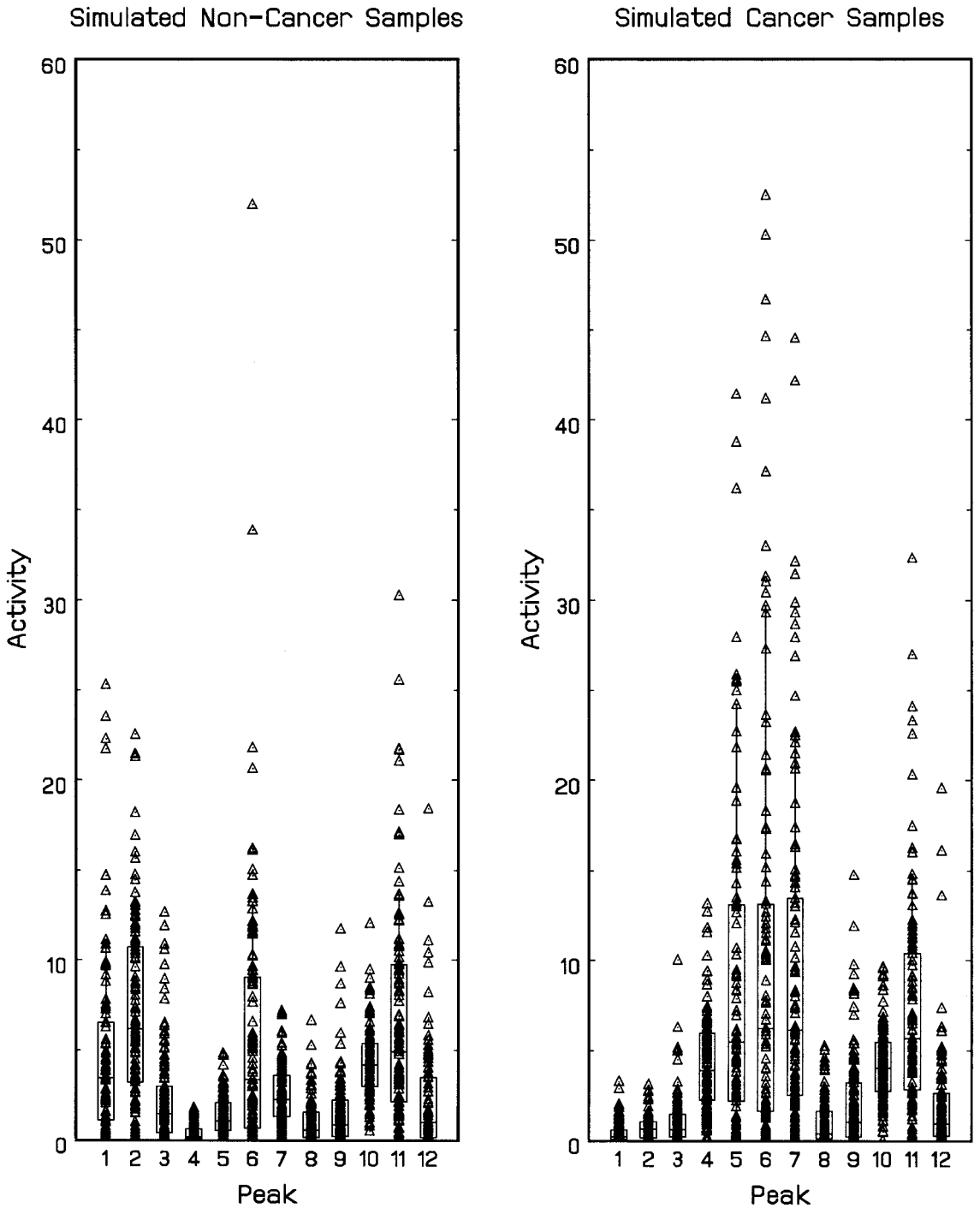

Fig. 1. A simulation of the boosted decision tree classifier.

Twelve peaks were assigned to each of the 200 simulated samples. *Peaks 1*, *2*, and *3* were higher in noncancer samples, whereas peaks *4*, *5*, *6*, and *7* were higher in cancer samples. Peaks *8*, *9*, *10*, *11*, and *12* were not informative. The selected peaks in six cycles of AdaBoost were *4*, *2*, *1*, *4*, *2*, and *7*. The classifier combined the six decision stumps with a resulting training error of 5% (10 of 200 were misclassified) and test error of 5.5% (11 of 200 were misclassified).

ing, and at the same time, the generalization error kept decreasing, finally reaching zero on round 265 and then staying at zero. For the boosted decision tree classifier distinguishing healthy men from men with BPH, after the training error reached zero (on round 9), the minimal margin kept increasing. The learning process did not stop when the training error became zero; on the contrary, the learning algorithm continued to enlarge the minimal margin between the two classes. Therefore, as long as the minimal margin keeps increasing, adding more base classifiers will not likely cause overfitting.

The first boosting classifier (AdaBoost Classifier) for

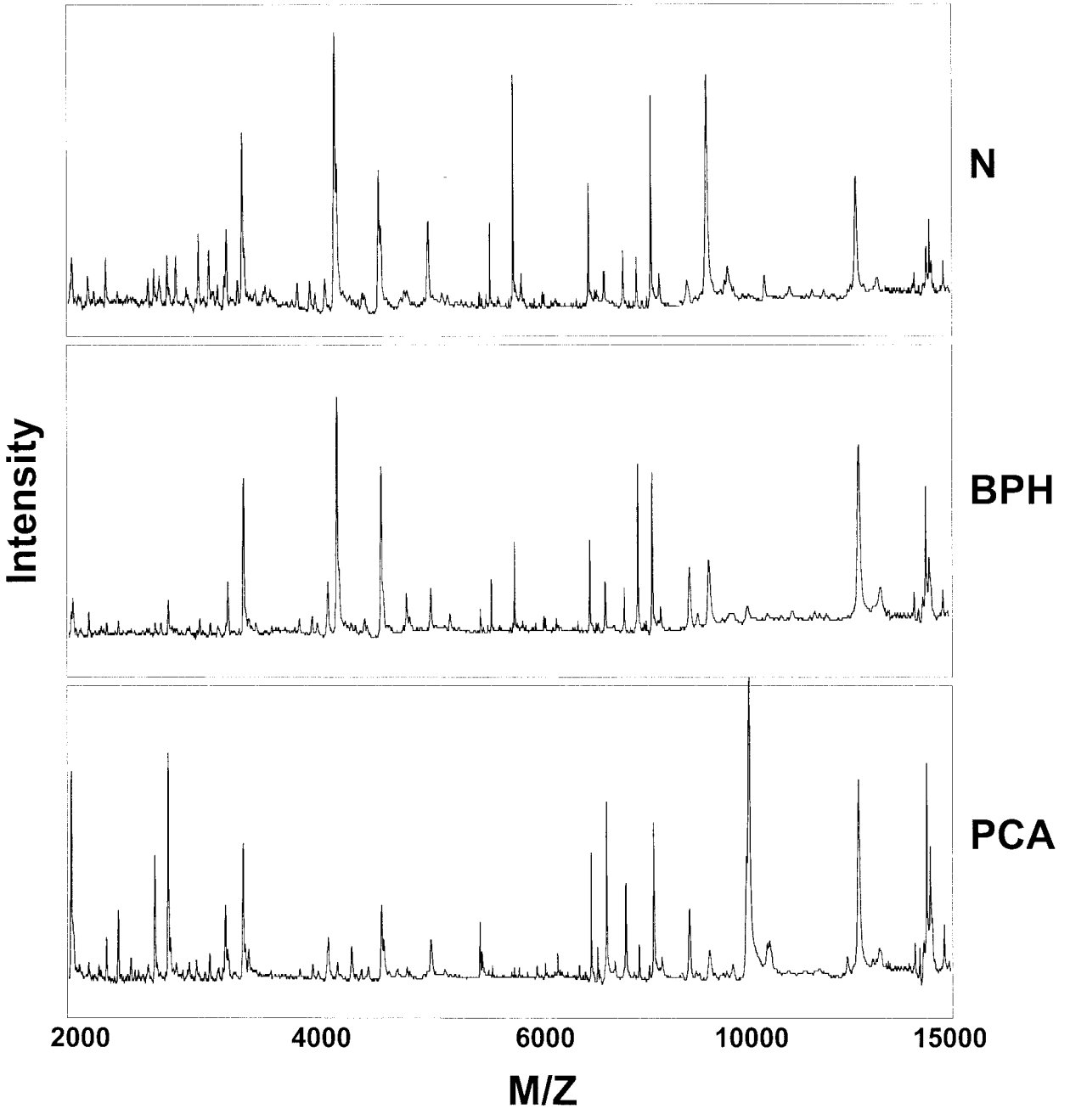

Fig. 2. Representative raw spectra of the peaks resolved between 2 and 15 kDa.

distinguishing noncancer from PCA consisted of 400 base classifiers, including 62 peaks, with a 0 error rate in both 326 training samples and in 60 test samples. When the number of base stumps (i.e., the number of rounds) was >47, the training error was zero, but the testing error (generalization error) was 0.0333. The generalization error slowly declined as the number of base stumps increased. After round 265, the generalization error remained zero. The 100 decision stumps for distinguishing healthy men from men with BPH also obtained a 0 error rate for both the 158 training and 30 test samples. In this case, the training error became zero on round 9 and the generalization error for 30 test samples was 0, beginning with round 1. When we combined these two boosted decision stumps, 100% separation was achieved for the three classes, healthy, BPH, and PCA, in both training and test sets (Table 1).

On the other hand, this classifier combined 500 base classifiers and 74 peaks. For the purpose of interpretation, there is a need to know which peaks are most important in distinguishing cancer from noncancer and which peaks are most important in distinguishing men with BPH from

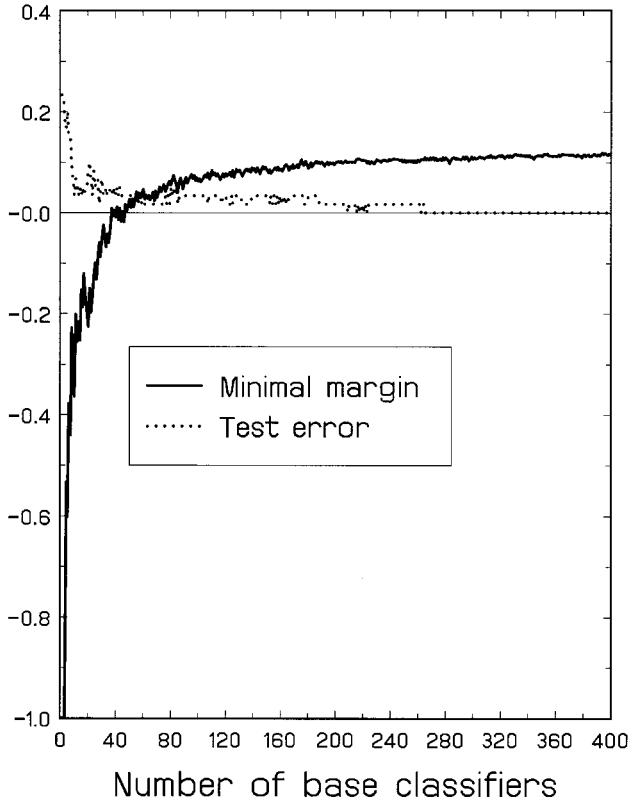

Fig. 3. The minimal margin and the generalization error curves for the boosted decision tree classifier.

Note that as the training error reaches zero (*round 47*), the minimal margin continues to increase and the generalization (test) error continues to decrease toward zero (*round 256*).

healthy men. We therefore needed a classifier with many fewer peaks. To construct this parsimonious classifier, we used the Boosted Decision Stump Feature Selection (BDSFS) algorithm *(10)*. This is essentially the same process as is used in the AdaBoost algorithm except that in each round only new features can be selected. In the case of the simulated data, the peaks selected in five cycles were 4, 2, 1, 7, and 5. For our PCA data, this classifier (BDSFS classifier) used 21 peaks, consisting of the 12 masses listed in Table 2, for distinguishing cancer from noncancer, and the first nine masses (Table 2) for distinguishing healthy men from men with BPH. This classifier achieved a sensitivity and specificity in the test set of 93.8%. In this case, the interpretation is much easier than the AdaBoost classifier, which contains 74 peaks (Table 1). However, the minimal margin for the BDSFS classifier is −0.2555, whereas the minimal margin for the AdaBoost classifier is 0.1143. Therefore, the AdaBoost classifier will be more accurate than the BDSFS classifier in discriminating PCA from the noncancer groups for new (unknown) samples.

We also applied a cross-validation approach to estimate the accuracy of the two classifiers in a process known as jack-knifing. We combined the training and test sets and randomly selected 10% of the samples from each of the three classes to be the test set, using the remaining 90% of the samples as a training set to construct an AdaBoost classifier with 500 base classifiers and a BDSFS classifier with 21 base classifiers. We repeated the procedure 10 times. Table 3 shows our results for each classifier. For the AdaBoost classifier, the estimated sensitivity in the test sets was 98.5% with a 95% confidence interval of 96.5–99.7%, and the specificity was estimated at 97.9% with a 95% confidence interval of 95.5–99.4%. For the BDSFS classifier, the sensitivity and specificity in the test sets were 91.1% (86.9–94.6%) and 94.3% (90.7–97.1%), respectively. The estimates of the sensitivity and specificity using the 10-fold cross-validation were in agreement with the estimates obtained from the original test set shown in Table 1. However, the confidence intervals with

Table 1. Classification of the training and test sets using the AdaBoost and the BDSFS algorithms.

	Training set				Test set			
	Total	Healthy	BPH	PCA	Total	Healthy	BPH	PCA
AdaBoost								
Healthy	82	82	0	0	15	15	0	0
BPH	77	0	77	0	15	0	15	0
PCA	167	0	0	167	30	0	0	30
Sensitivity		100.0%				100.0%		
Specificity		100.0%				100.0%		
Number of base classifiers: 500								
Minimal margin: 0.1143								
BDSFS								
Healthy	82	82	0	0	15	14	0	1
BPH	77	0	74	3	15	0	15	0
PCA	167	7	0	160	30	0	1	29
Sensitivity		95.8%				96.7%		
Specificity		98.1%				96.7%		
Number of base classifiers: 21								
Minimal margin: −0.2555								

Table 2. Peak masses[a] used by the BDSFS classifier.

Noncancer vs cancer		Healthy vs BPH	
Peak no.	Mass, Da	Peak no.	Mass, Da
1	9655.75	1	7819.75
2	9719.99	2	4579.73
3	6541.82	3	7844.00
4	6797.02	4	4071.18
5	6949.22	5	7054.17
6	7024.02	6	5297.55
7	8066.95	7	3486.21
8	8355.56	8	6099.08
9	3963.18	9	8943.08
10	4079.48		
11	7884.72		
12	6990.63		

[a] The peaks are listed in the order of their selection.

the cross-validation method were narrower than those for the original test set (because the sample size in the original test set was smaller).

Discussion

SELDI mass spectrometry using a protein chip that captures proteins based on their ability to selectively bind to a chemically activated copper surface through histidine, tryptophan, cysteine, or phosphorylated amino acids was capable of resolving an average of 80 serum proteins/peptides, ranging from 2 to 40 kDa. This is far less than the hundreds to thousands of proteins capable of being separated by two-dimensional electrophoresis, but the advantage over two-dimensional electrophoresis is the ability of SELDI to effectively resolve polypeptides and peptides smaller than 20 kDa. This has opened the door to readily resolve and study such peptides as potential biomarkers for diagnosis, prognosis, and therapeutic targets. Interestingly, the 21 proteins identified in this study to be the most critical for separating PCA from the noncancer groups are between 3 and 10 kDa. Since the introduction of SELDI, there has been a concern that such "peptides" might represent nonspecific degradation products of larger proteins. If this were true, we would not have been able to achieve high reproducibility in protein patterns (0.05% CV for peak location).

Even with only an average of 80 peaks per spectrum obtained between 2 and 40 kDa, there is still extremely high dimensionality of the data. Initial analysis of the 772 serum samples (386 run in duplicate) produced 63 157 peaks, which were reduced to 779 peaks after cluster analysis and peak alignment. Of the 779 peaks, 124 peaks were statistically found to have the highest potential to discriminate the three groups: healthy vs PCA, healthy vs BPH; and BPH vs PCA. Subsequent analysis of the 124 peaks in each of the 772 samples led to processing of >95 000 data points to identify the pattern or combination of masses that separates PCA from the noncancer groups. Because of this high dimensionality, only an artificial intelligent algorithm would be capable of analyzing such high volume of data to develop an efficient and reproducible classifier. We have evaluated several different models, including biostatistical *(11, 12)*, genetic clustering, and support vector machine algorithms. Although most could obtain 83–90% accuracy in differentiating PCA from the noncancer (BPH/healthy) groups, the decision tree model *(13)* was selected because it is easier to interpret than "black box" classifiers such as neural networks and biostatistical algorithms. Using the same data set described in the present study, we developed a single decision tree

Table 3. Classification of the training and test sets in 10-fold-stratified cross-validation.[a]

	Prediction							
	Training set				Test set			
	Total[a]	Healthy	BPH	PCA	Total	Healthy	BPH	PCA
AdaBoost								
Healthy	870	963	0	7	100	99	1	0
BPH	830	0	930	0	90	0	87	3
PCA	1770	2	0	1768	200	1	1	198
Sensitivity		99.9%				99.0%		
Specificity		99.6%				98.4%		
Number of base classifiers: 500								
BDSFS								
Healthy	870	837	6	27	100	92	2	6
BPH	830	1	789	40	90	1	85	4
PCA	1770	105	0	1665	200	9	8	183
Sensitivity		94.1%				91.5%		
Specificity		96.1%				94.7%		
Number of base classifiers: 21								

[a] In each round of cross-validation, 90% of each group were assigned to be the training set and the remaining 10% were held out as the test set. Specifically, there were 87 healthy, 83 BPH, and 177 PCA samples in the training set and 10 healthy, 9 BPH, and 20 PCA samples in the test set. The total sample size is the sum in the 10 times cross-validation.

base classifier with nine masses between 2 and 10 kDa that achieved a sensitivity of 83% and a specificity of 97% for differentiating PCA from the noncancer groups (i.e., BPH and healthy) *(14)*. However, a single decision tree classifier's predictive power may not be as good as other learning algorithms, such as neural networks and support vector machine. Furthermore, assays for the early detection of cancer need to be highly accurate to avoid generating too many false positives. The present study was initiated to determine whether we could increase the predictive power of the decision tree classifier. Tremendous improvement in the predictive power of decision tree classifiers has been reported recently by use of voting methods, such as boosting *(15, 16)*, and bootstrap methods. In one voting method, called the bagging method *(17, 18)*, the decision tree model is fitted many times on randomly resampled observations (bootstrap subsamples) and then combines the decision trees using simple voting. Another approach is the boosting method *(7)*, referred to as the AdaBoost algorithm, which fits the learning algorithm (such as the decision tree model) many times on weighted observations and then combines the decision trees by use of weighted voting. In both bagging and boosting, the combined classifier has better performance than each of the individual base decision trees in the test set. We chose the boosting approach over the bagging algorithm because it is generally more accurate in the test samples than the bagging approach *(17)*. With the AdaBoost algorithm, we established a classifier that was error free in predicting, for both the training and blinded test sets, whether the sample was from a patient diagnosed with PCA or BPH or from a healthy donor. Although this classifier produced high sensitivity and specificity, it used 74 protein mass values (peaks) and required combining 500 base decision tree classifiers, making it highly accurate but difficult to interpret. Other models, such as wavelets analysis *(11)* and support vector machines (G.L. Wright, unpublished data), can reach similar high accuracy but with the same difficulty, especially in identifying the protein masses used in the classifier. The BDSFS classifier with 21 peaks selected is much easier to interpret. It was slightly less accurate than the AdaBoost classifier; it misclassified 1 of 15 BPH samples as PCA and 1 of 29 PCA samples as healthy; whereas all 14 samples from healthy, unaffected men were correctly identified. This parsimonious classifier still achieved a respectable 93.8% for both sensitivity and specificity, using 21 peaks only. The 21 peaks identified by this algorithm are potential biomarkers that will be tested in future studies.

The PSA test is the current screening test for PCA, and if positive, biopsies are obtained from each lobe of the prostate. Many consider this test the best for any human cancer, but it is far from being a perfect test for early detection of PCA. Although it has a high sensitivity of >90%, its specificity is only 25% in distinguishing PCA from BPH; and some men with PCA have PSA concentrations within reference values *(1)*. Because of the low specificity, men are subjected to unnecessary biopsies, causing considerable anxiety when they in fact do not have cancer. Current evidence also suggests that a preoperative serum PSA $<10\ \mu g/L$ is not a useful biomarker for predicting disease presence, volume, grade, or rate of postoperative failure *(1)*. On the basis of these facts, there is a need for better biological markers than PSA and all its molecular forms can provide. Provided that the accuracy of the boosting decision tree classifier can be validated on a larger number of samples and evaluated at multiple sites, including testing the validity of the profiling assay with samples from noncancer patients, SELDI protein profiling combined with a bioinformatics classifier may provide that "better" test for the early detection and diagnosis of PCA. Support for this potential appears in reports from other investigators who have achieved similar results for ovarian and breast cancer, using SELDI combined with a bioinformatics classifier different from the classification system used in the present study *(19, 20)*. Overall, these initial studies suggest that SELDI provides a unique opportunity to develop an innovative proteomic approach for cancer diagnosis.

The identity of the peak masses used in the classifier is not necessary for making a diagnosis. The only requirement for this classification system to make an accurate diagnosis is that the biomarkers be reproducibly detected by SELDI and accurately selected by the classifier. Obtaining a name for each of the masses used in the classifier will not make the classification system better or more accurate. Knowing the identity of the protein biomarkers is, however, essential from a discovery perspective. The identities of the peptide/protein biomarkers will be needed to understand the biological role these proteins have in the oncogenesis of PCA. Such information could lead to better therapeutic interventions. Knowing the identities will facilitate the production of antigen and antibody reagents for development of classic multiplex immunoassays and antibody arrays, should the profiling approach fail to be developed into a clinical assay. For these reasons, protein identification of the potential biomarkers is in progress.

In conclusion, the high sensitivity and specificity achieved by the combined use of multiple serum biomarkers provides supporting evidence that SELDI, combined with a learning algorithm, not only can facilitate the discovery of new and better biomarkers for PCA, but has potential for being developed into a novel clinical diagnostic assay.

This study was supported by grants from the National Cancer Institute Early Detection Research Network (CA 85067), the Department of Defense (DAMD17-02-1-0054), and the Virginia Prostate Center.

References

1. Stamey TA, Johnstone IM, McNeal JE, Lu AY, Yemoto CM. Preoperative serum prostate specific antigen levels between 2 and 22 ng/ml correlate poorly with post-radical prostatectomy cancer morphology: prostate specific antigen cure rates appear constant between 2 and 9 ng/ml. J Urol 2002;167:103–11.
2. Merchant M, Weinberger SR. Recent advancements in surface-enhanced laser desorption/ionization-time of flight-mass spectrometry. Electrophoresis 2000;21:1164–77.
3. Kuwata H, Yip TT, Yip CL, Tomita M, Hutchens TW. Bactericidal domain of lactoferrin: detection, quantitation and characterization of lactoferricin in serum by SELDI affinity mass spectrometry. Biochem Biophys Res Comm 1998;245:764–73.
4. Wright GL Jr, Cazares LH, Leung S-M, Nasim S, Adam BL, Yip TT, et al. ProteinChip surface enhanced laser desorption/ionization (SELDI mass spectrometry: a novel proteomic technology for detection of prostate cancer biomarkers in complex protein mixtures. Prostate Cancer Prostate Dis 1999;2:264–7.
5. Vlahou A, Schellhammer PF, Mendrinos S, Patel K, Kondylis FI, Gong L, et al. Development of a novel proteomic approach for the detection of transitional cell carcinoma of the bladder in urine. Am J Pathol 2001;158:1491–502.
6. Pepe MS. Receiver operating characteristic methodology. J Am Stat Assoc 2000;95:308–11.
7. Freund Y, Schapire R. A decision-theoretical generalization of on-line learning and an application to boosting. J Computer Syst Sci 1997;55:119–39.
8. Hastie T, Tibshirani R, Friedman J. The elements of statistical learning. New York: Springer-Verlag, 2001:301pp.
9. Friedman J, Hastie T, Tibshirani R. Additive logistic regression: a statistical view of boosting. Ann Stat 2000;28:337–407.
10. Das S. Filters, wrappers and a boosting-based hybrid for feature selection. In: Brodley CE, Danyluk AP, eds. Proceeding of the Eighteenth International Conferences on Machine Learning. San Francisco, CA: Morgan Kaufmann Publishers, 2001:74–81.
11. Qu Y, Adam B-L, Thornquist M, Potter J, Yasui Y, Davis J, et al. Data reduction using discrete wavelet transform in discriminant analysis with very high dimension. Biometrics 2002;in press.
12. Yasui Y, Pepe M, Thompson ML, Adam BL, Wright G Jr, Qu Y, et al. A data-analytic strategy for protein-biomarker discovery: profiling of high-dimensional proteomic data for cancer detection. Technical Report No. 177. Seattle, WA: University of Washington, Department of Biostatistics, 2001.
13. Breiman L, Friedman JH, Olsen RA, Stone CJ. Classification and regression trees. Belmont, CA: Wadsworth International Group, 1984:203–15.
14. Adam B-L, Qu Y, Davis JW, Ward MD, Clements MA, Cazares LH et al. Serum protein fingerprinting coupled with a pattern-matching algorithm distinguishes prostate cancer from benign prostate hyperplasia and healthy men. Cancer Res 2002;62:3609–14.
15. Drucker H, Cortes C. Boosting decision trees. Adv Neural Inf Process Syst 1996;8:479–85.
16. Schapire R, Freund Y. Boosting the margin: a new explanation for the electiveness of voting methods. Ann Stat 1998;26:1651–86.
17. Breiman L. Bagging predictors. Machine Learning 1996;26:123–40.
18. Breiman L. Statistical modeling: the two cultures. Stat Sci 2001; 16:199–231.
19. Petricoin EF III, Ardkani AM, Hitt BA, Levine PJ, Fusaro VA, Steinberg SM, et al. Use of proteomic patterns in serum to identify ovarian cancer. Lancet 2002;359:572–7.
20. Li J, Zhang Z, Rosenzweig J, Wang YY, Chan D. Potential serum biomarkers identified by SELDI mass spectrometry can discriminate breast cancer from non-cancer patients [Abstract]. Proc Am Assoc Cancer Res 2002;43:136.